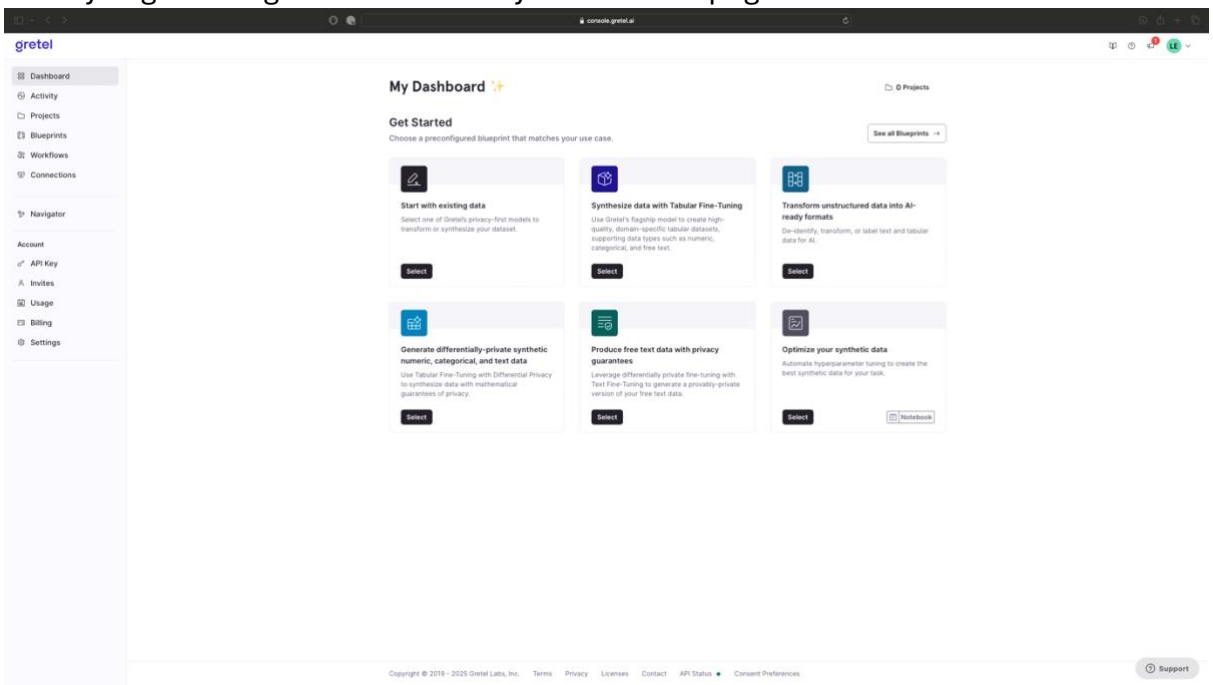
This document will provide a quick overview of using Gretel.ai for synthetic data generation.

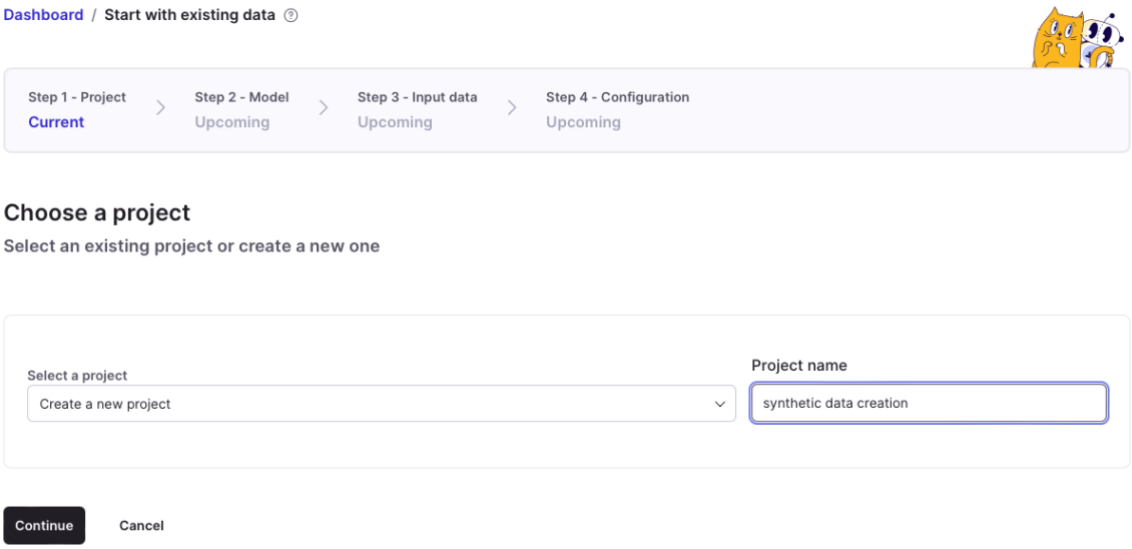
First off, go to this link. https://gretel.ai

It asks for a work email address but I had success signing in with my personal gmail account and my deakin student email address.

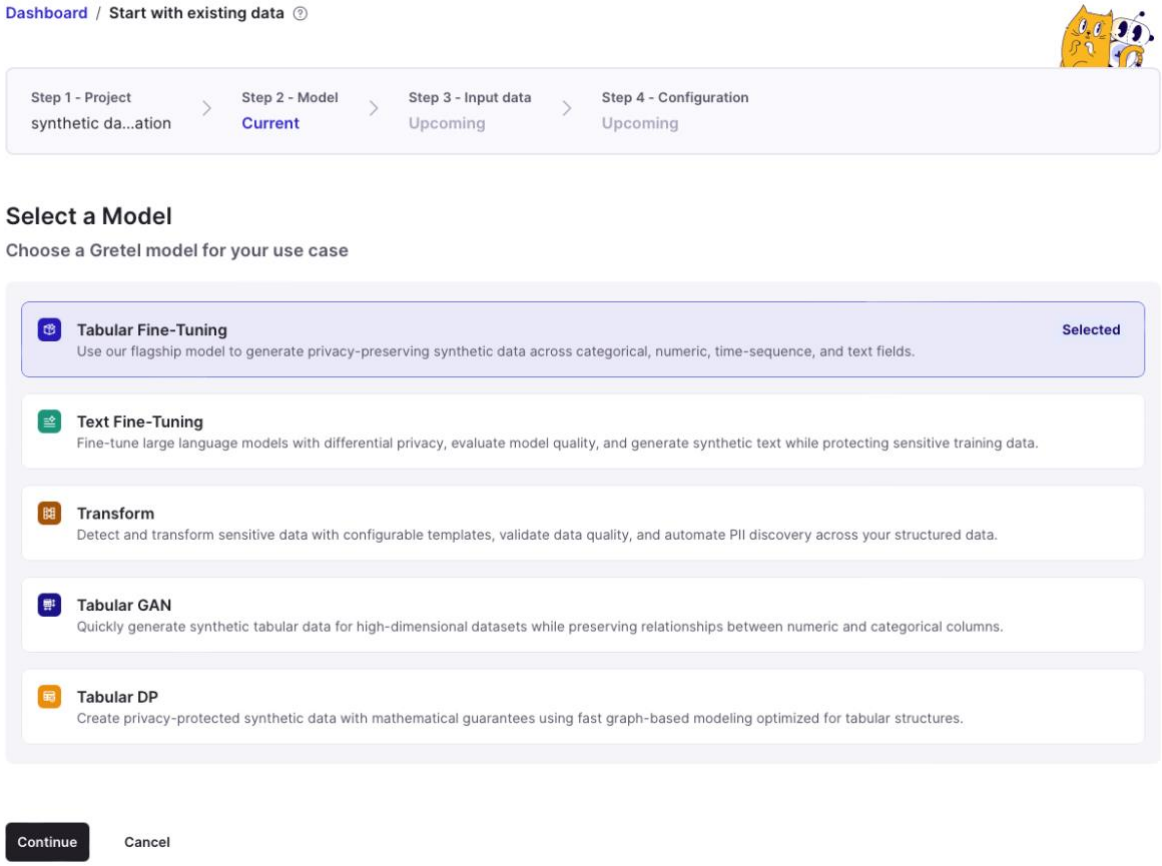
Once you get through the verification you'll see this page.



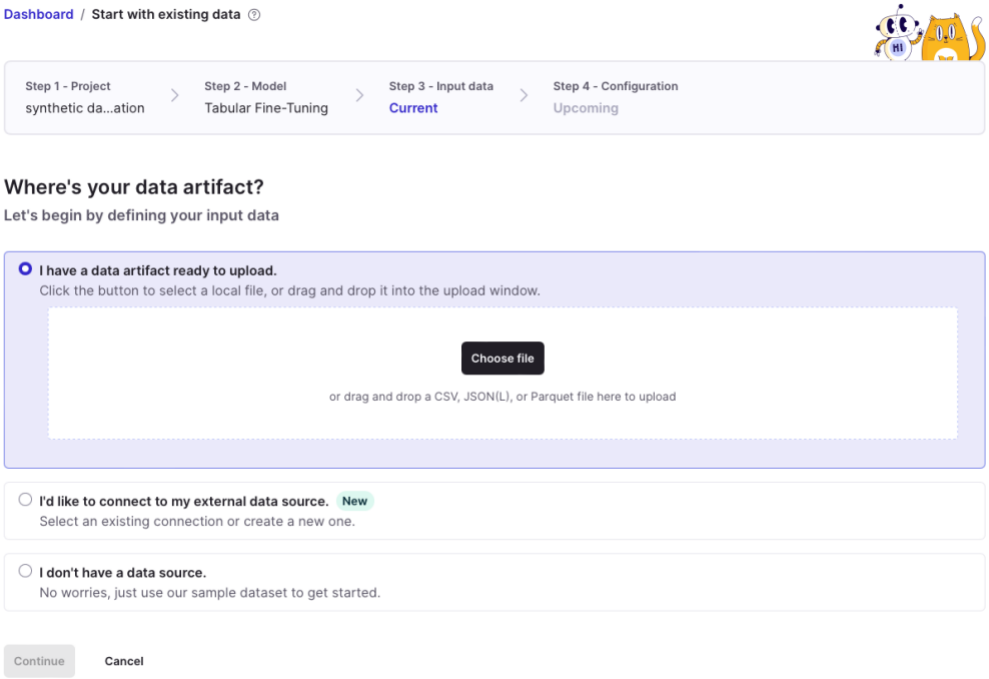
Click on Start with Existing data and you'll be prompted to create a project.



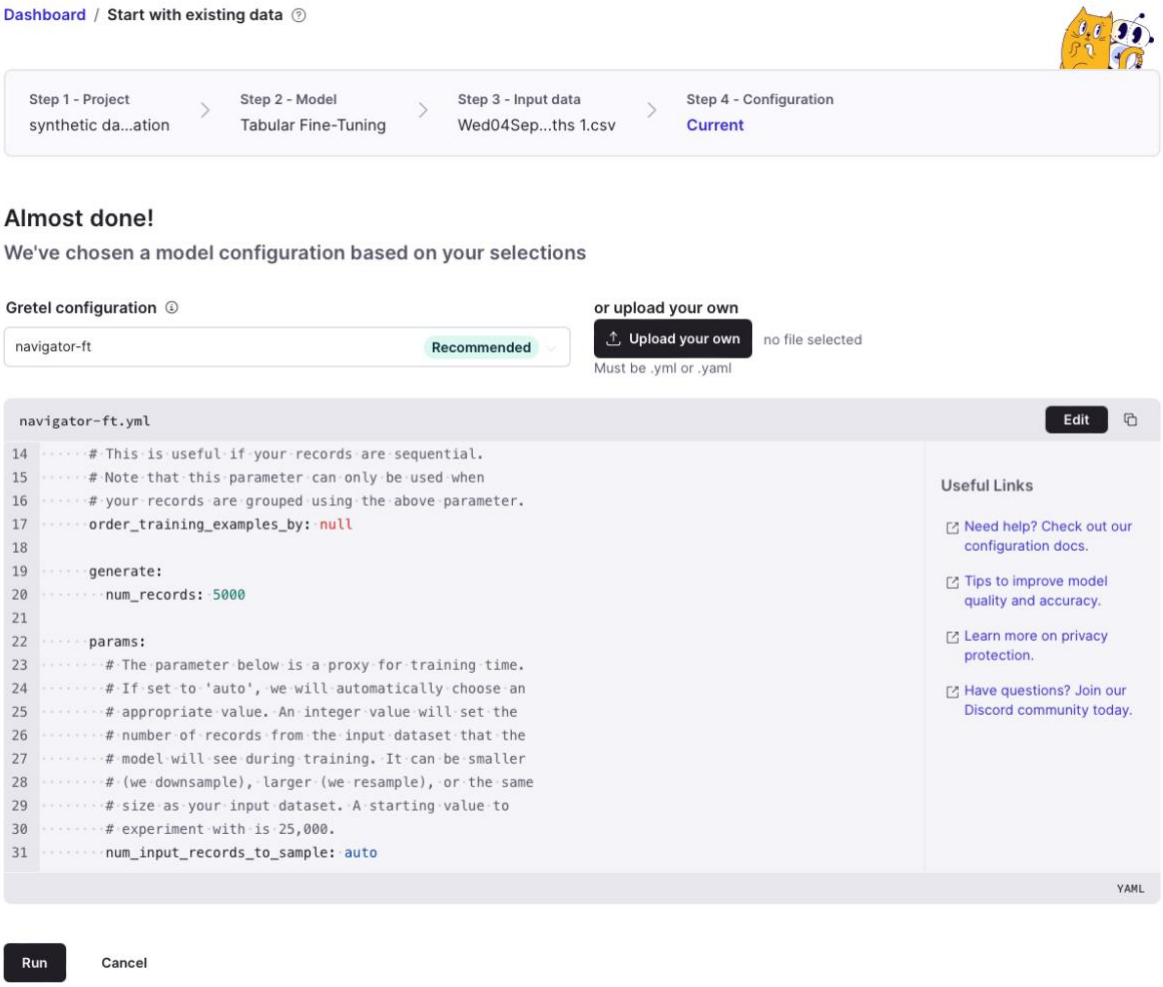
Select your model. Have a read of the descriptions but the top one that is preselected appears to be the most relevant.



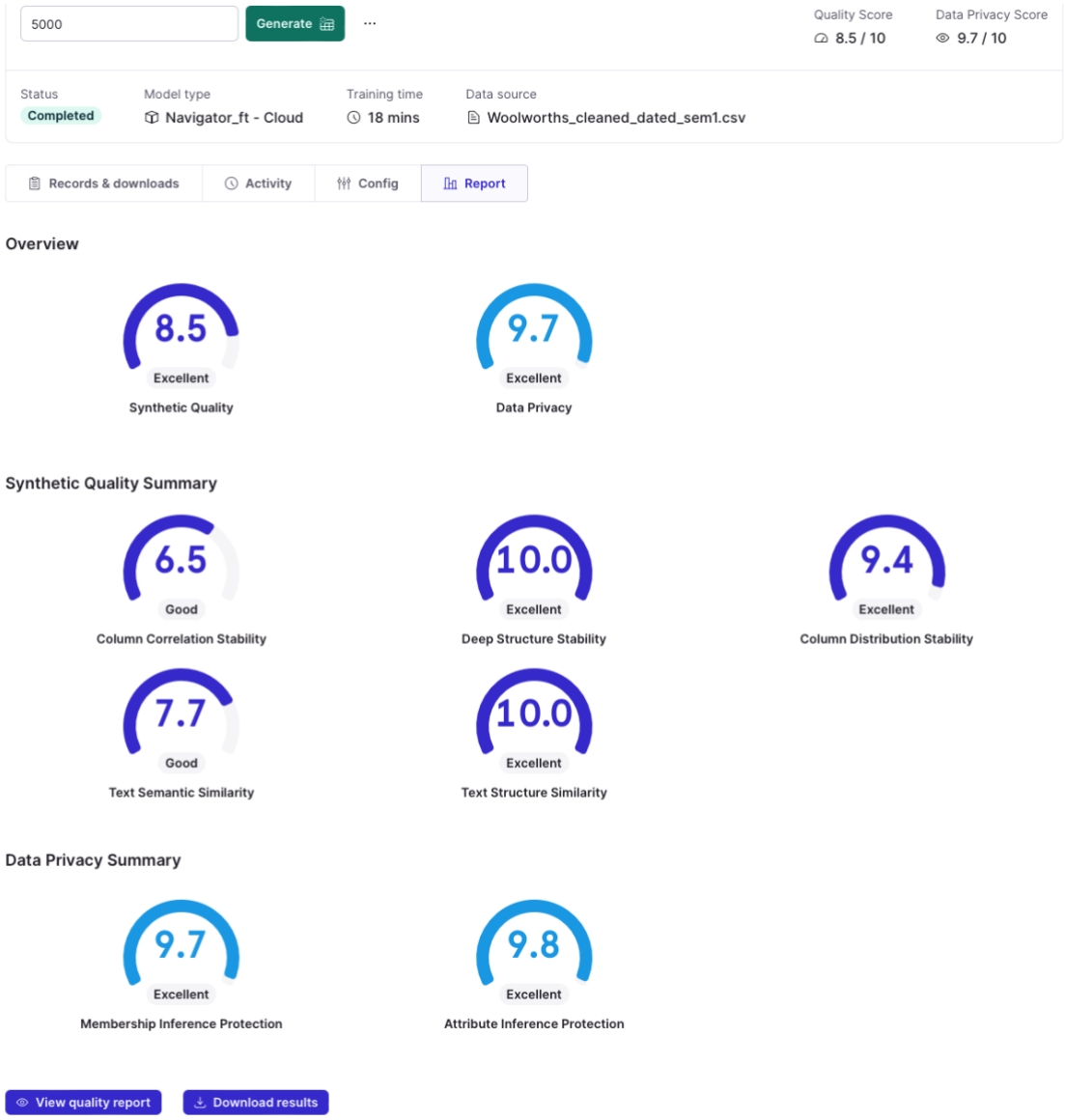
The next page is where you drop in your input data that you want the model to be trained on. I think this will be extremely useful for creating transactional data. Just make sure we're covering the categories agreed on in the meeting last night.



You'll then be at the final page before running the process. It will be automatically set to generate 5000 records, if you want to change this scroll down and edit the yml file.



Once you start the process it can take a while to complete. Some of my attempts took between 10 and 20 minutes. (Depends on amount of data you use)
You'll receive a report on the quality of the data along with the download button for the data you just generated.



It is possible to do this through terminal using the CLI and the Gretel API key which is available, but the web process is much more straightforward for the same results.

Here is a link to documentation. https://docs.gretel.ai

In my experimenting I combined the Woolworths csv files into one dataframe and added a date column. Gretel did not come up with new dates but instead generated for those dates. So, if you were to work on grocery price generation another approach would be needed.

Making me lean to focusing on using it for transactional synthetic data.